POPGENE VERSION 1.31

Microsoft Window-based Freeware for
Population Genetic Analysis

Quick User Guide

A joint Project Development by
Francis C. Yeh and Rong-cai Yang, University of Alberta
And
Tim Boyle, Centre for International Forestry Research

August 1999

New Feature in Current Version

For Windows 95, 08 and NT Users (32-bit version)

This version (1.31) has a new graphics interface that produces publication quality dendrograms.

POPGENE generates two dendrograms for each populations analysis, base on Nei's regular and unbiased genetic distance measures. Immediately after each of these dendrograms in the output file, you will now see "**File Name: dgram1.plt** " or "**File Name: dgram2.plt** ". These two files are stored in the same directory as your output file. They are the files you use for printing publication quality dendrogram.

Now, open your word-processing package such as Microsoft Word or Corel WordPerfect. Use the statements/commands "insert → picture → file from" to bring these files, "**dgram1.plt**" or "**dgram2.plt"** into your word processor.

For Windows 3.11 Users (16-bit version)

Users must first download and install the following file:

ftp://ftp.microsoft.com/Softlib/MSLFILES/HPGL.EXE

Then, follow the procedures detailed above for Windows 95, 98 and NT.

Installing POPGENE from diskette/CD

1. Start Microsoft Windows (Windows 3.11, Windows 95, 98 and NT).

2. Insert the POPGENE Disk into your floppy drive/CD Drive.

3. Run A: POPGEN16.EXE (for 16-bit operating environment under Windows 3.11) or A: POPGEN32 (for 16-bit operating environment under Windows 95, 98 and NT) where ‘A’ is the floppy drive number or CD drive number on your computer.

4. Follow the instruction on your screen.

5. Go to the directory where you installed POPGENE and double click the POPGENE icon to start the program.

Installing POPGENE that you downloaded from website

1. Start Microsoft® Window (Windows 3.11, Windows 95, 98 and NT).

2. Go to the directory where you stored your downloaded POPGENE.

3. Type or double click on the file you saved.

4. Follow the instructions on your computer screen.

5. Go to the directory where you have installed POPGENE and double click the POPGENE icon to start the program.

Uuinstall POPGENE ?

32- bit version

Start Microsoft® Windows and select "Start" menu, "Settings", then "Control Panel".

Select "Add/Remove Programs" and double click on "Population Genetics Analysis ".

Follow the instructions on your computer screen.

16- bit version

Start Microsoft® Windows and select PopGen16 folder.

Double click on "Uninstall PopGen16".

Follow the instructions on your computer screen.

Introduction

POPGENE is a user-friendly **Microsoft**® Window-based computer package for the analysis of genetic variation among and within natural populations using co-dominant and dominant markers and quantitative traits. This package provides the Windows graphical user interface that makes population genetics analysis more accessible for the casual computer user and more convenient for the experienced computer user. Simple menus and dialog box selections enable you to perform complex analysis and produce scientifically sound statistics, thereby assisting you to adequately analyze population genetic structure using the target markers/traits.

The current version (version 1.30) is designed specifically for the analysis of co-dominant and dominant markers using haploid and diploid data. It performs most types of data analysis encountered in population genetics and related fields. It can be used to compute summary statistics (*e.g.*, allele frequency, gene diversity, genetic distance, *F*-statistics, multilocus structure, *etc.*) for (1) single-locus, single populations; (2) single-locus, multiple populations; (3) multilocus, single populations and (4) multilocus, multiple populations.

Version 1.30 also includes the module for quantitative traits. However, we are not supporting it at this time because of the large RAM requirement when analyzing quantitative genetic data.

Hardware and Software Requirements

To be able to run **POPGENE**, your hardware and software must meet the following minimum requirements:

Hardware*: An IBM PC or compatible with an Intel 386 processor or higher CPU and 8 or more MB of random access memory (RAM). A mathematical coprocessor is required to achieve a reasonable computing speed.

Software: Windows 3.11 (16-bit version) or later version, Windows 95, 98 or NT (32-bit version)

APPLE computer users can run POPGENE on PowerPc or the new G3, but must first install a software such as "Virtual PC" or “Soft Windows”. POPGENE runs effortlessly under "Virtual PC" and “Soft Windows”.

* Neutrality tests and all multilocus genetic estimates are computationally demanding. We strongly recommend running **POPGENE** on a Pentium-based PC with 16 or more MB of RAM.

Overview of POPGENE Windows

This package is written in **Borland**® C++ 4.51, a powerful professional tool for creating Windows applications using the C and C++ languages.

The **POPGENE** Windows computing environment consists of two types of windows: **Data display windows** and **Dialog boxes**.

POPGENE is menu driven. Most features are accessed by making selections from the menus. The **main menu bar** contains eight menus:

File. Use the File menu to create a new data file or open an existing file.

Edit. Use the Edit menu to modify or copy text from other Windows.

Search. Use the Search menu to find and/or replace selected text.

Co-Dominant. Use the Co-Dominant menu to invoke population genetics analysis using co-dominant markers.

Dominant. Use the Dominant menu to invoke population genetics analysis using dominant markers.

Quantitative. Use the Quantitative menu to invoke population genetics analysis using quantitative traits.

Window Use the Window menu to arrange, select and control the attributes of different windows.

Help. Use the Help menu to open a standard Microsoft Help window containing information on how to use different features of the POPGENE package.

At the top of the data window, there is a **Tool Bar** to provide quick, easy access to the special features of the window.

At the bottom of the POPGENE application window, there is a **Status Bar** to indicate the current status of the window, including the cursor position and the size of the input data file (in bytes).

Overview of POPGENE Computing Programs

Below is a brief description of each of the programs developed for POPGENE/Co-Dominant and Dominant markers. For more detailed discussion on the algorithms from which the programs have been developed, you should consult with the original references cited at the end of this section. These references will be also helpful in assisting you to interpret the outputs from the programs.

POPGENE/Co-Dominant and Dominant markers have two dialog boxes: **Haploid Data Analysis** and **Diploid Data Analysis**. In each of these boxes, there are three levels of **Hierarchical Structure** given as three check boxes: **Single Populations**, **Groups** and **Multiple populations**. Estimation of **Single Locus** and **Multilocus** genetic parameters is carried out by clicking one or more **Hierarchical Structure** check boxes and one or more **Single Locus** and **Multilocus** check boxes.

HAPLOID DATA ANALYSIS

Gene Frequency: Estimates gene frequencies at each locus from raw data. Missing values are excluded from such estimation.

Allele Number: Counts the number of alleles with nonzero frequency.

Effective Allele Number: Estimates the reciprocal of homozygosity (Hartl and Clark 1989, p.125).

Polymorphic Loci: Percentage of all loci that are polymorphic regardless of allele frequencies.

Gene Diversity: Estimates Nei's (1973) gene diversity.

Shannon Index: Estimates Shannon's information index as a measure of gene diversity.

Homogeneity Test: Constructs two-way contingency tables and carries out chi-square (χ^2) and likelihood ratio (G^2) tests for homogeneity of gene frequencies across populations. The tests are carried out for **Groups** or **Multiple Populations**.

F-Statistics: Estimates Nei's (1973) G_{ST} for **Groups** or **Multiple Populations**, and estimates both G_{ST} and G_{CS} for **Groups** and **Multiple Populations**.

Gene Flow: Estimates gene flow from the estimate of G_{ST} or F_{ST} (Slatkin and Barton 1989). The estimation is made for **Groups** or **Multiple Populations**.

Genetic Distance: Estimates Nei's (1972) genetic identity and genetic distance and Nei's (1978) unbiased genetic identity and genetic distance. The estimation is made for **Groups** or **Multiple Populations**.

Dendrogram: Draws a dendrogram based on Nei's genetic distances using UPGMA. This program is an adoption of program NEIGHBOR of PHYLIP version 3.5c by Joe Felsenstein. The drawing is executed for **Groups** or **Multiple Populations**.

Neutrality Test: Performs the Ewens-Watterson test for neutrality using the algorithm given in Manly (1985).

Two-locus LD: Estimates gametic disequilibria between pairs of loci and χ^2 tests for significance (Weir 1979) for **Single Populations**, and performs Ohta's (1982a, b) two-locus analysis of population subdivision (D-Statistics) for **Multiple Populations**.

Brown: Compute observed and expected moments of K, the number of heterozygous loci between two randomly chosen gametes in a population, as well as multilocus indices and 95% confidence limits from these moments (Brown et al., 1980) for **Single Populations**, and partition the total and average variances of K in a mixed pool of several populations into single-locus and two-locus components (Brown and Feldman 1981) for **Multiple Populations**.

Smouse: Codes the most frequent allele as one (1) and a "synthetic" allele consisting of all the other alleles combined as zero (0) (Yang and Yeh 1993). Estimates average interlocus correlation based on the coded data in a population (Smouse and Neel 1977) for **Single Populations** and estimates among- and within-population interlocus correlations for **Multiple Populations**.

DIPLOID DATA ANALYSIS

Genotypic Frequency: Estimates genotypic frequencies observed at each locus from raw data **only for co-dominant markers**. Missing values are excluded from such estimation.

HW Test: Computes expected genotypic frequencies under random mating using the algorithm by Levene (1949), and perform chi-square (χ^2) and likelihood ratio (G^2) tests for Hardy-Weinberg equilibrium at each locus **only for co-dominant markers**.

Fixation Index: Estimates F_{IS} as a measure of heterozygote deficiency or excess (Wright 1978) **only for co-dominant markers**.

Allele Frequency: Estimates gene frequencies at each locus from raw data. Missing values are excluded from such estimation.

Allele Number: Counts the number of alleles with nonzero frequency.

Effective Allele Number: Estimates the reciprocal of homozygosity (Hartl and Clark 1989, p.125).

Polymorphic Loci: Percentage of all loci that are polymorphic regardless of allele frequencies.

Obs. Homozygosity: Estimates proportion of observed homozygotes at a given locus.

Exp. Homozygosity: Estimates proportion of expected homozygotes under random mating (see **Exp. Heterozygosity** for appropriate references).

Obs. Heterozygosity: Estimates proportion of observed heterozygotes at a given locus **only for co-dominant markers**.

Exp. Heterozygosity: Estimates proportion of expected heterozygotes under random mating **only for co-dominant markers**. Two estimates are given. The first is Nei's (1973) heterozygosity. The second the expected heterozygosity estimated using the algorithm of Levene (1949), which is the same as Nei's (1978) unbiased heterozygosity.

Shannon Index: Estimates Shannon's (1949) information index as a measure of gene diversity.

Homogeneity Test: Constructs two-way contingency tables and carries out chi-square (χ^2) and likelihood ratio (G^2) tests for homogeneity of gene frequencies across populations. The tests are carried out for **Groups** or **Multiple Populations**.

F-Statistics: Estimates F-statistics (F_{IT}, F_{ST} and F_{IS}) for **Groups** or **Multiple Populations** (Hartl and Clark 1989), but estimates F-statistics for a three-level sampling hierarchy in random populations using a quite different approach by Weir (1990) for both **Groups** and **Multiple Populations**.

Gene Flow: Estimates gene flow from the estimate of G_{ST} or F_{ST} (Slatkin and Barton 1989). The estimation is made for **Groups** or **Multiple Populations**.

Genetic Distance: Estimates Nei's (1972) genetic identity and genetic distance and Nei's (1978) unbiased genetic identity and genetic distance. The estimation is made for **Groups** or **Multiple Populations**.

Dendrogram: Draws a dendrogram based on Nei's genetic distances using UPGMA. This program is an adoption of program NEIGHBOR of PHYLIP version 3.5c by Joe Felsenstein. The drawing is executed for **Groups** or **Multiple Populations**.

Neutrality Test: Performs the Ewens-Watterson test for neutrality using the algorithm given in Manly (1985).

Two-locus LD: Estimates Burrows' composite measure of linkage disequilibria between pairs of loci and (χ^2) tests for significance (Weir 1979) for **Single Populations**, and performs Ohta's (1982a, b) two-locus analysis of population subdivision (D-Statistics) for **Multiple Populations**. **Only for co-dominant markers**.

Smouse: Codes a homozygote for the most frequent allele as one (1), a homozygote for the "synthetic" allele consisting of all the other alleles combined as zero (0) and their heterozygote as one-half (1/2). Estimates average interlocus correlation based on the coded data in a population and test for both Hardy-Weinberg and linkage disequilibria (Smouse et al. 1983) for **Single Populations** and estimates among- and within-population interlocus correlations for **Multiple Populations**. In **Multiple Populations** case, no attempt was given to estimate Hardy-Weinberg disequilibrium in a subdivided population as these estimates would be equivalent to F-statistics given above. **Only for co-dominant markers**.

References

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure in natural populations of *Hordeum spontaneum*. Genetics 96:523-536.

Brown AHD, Feldman MW (1981) Population structure of multilocus associations. Proc Natl Acad Sci USA 78:5913-5916.

Hartl DL, Clark AG (1989) Principles of population genetics. 2nd ed. Sinauer Associates, Sunderland, MA.

Levene H (1949) On a matching problem in genetics. Ann Math Stat 20:91-94.

Manly BFJ (1985) The statistics of natural selection. Chapman and Hall, London.

Nei M (1972) Genetic distance between populations. Am Nat 106:283-292.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321-3323.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Ohta T (1982a) Linkage disequilibrium with the island model. Genetics 101:139-155.

Ohta T (1982b) Linkage disequilibrium due to random drift in finite subdivided populations. Proc Natl Acad Sci USA 79:1940-1944.

Shannon CE, Weaver W (1949) The mathematical theory of communication. Univ. of Illinois Press, Urbana.

Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. Evolution 43:1349-1368.

Smouse PE, Neel JV (1977) Multivariate analysis of gametic disequilibrium in the Yanomama. Genetics 85:733-752.

Smouse PE, Neel JV, Liu W (1983) Multiple-locus departures from panmictic equilibrium within and between village gene pools of Amerindian Tribes at different stages of agglomeration. Genetics 104:133-153.

Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35:235-254.

Weir BS (1990) Genetic Data Analysis. Sinauer Associates, Sunderland, MA.

Wright S (1978) Variability within and among natural populations. Vol. 4. The Univ. Of Chicago Press, Chicago.

Yang R-C, Yeh FC (1993) Multilocus structure in *Pinus contorta* Dougl. Theor Appl Genet 87:568-576.

What Can POPGENE Do for Your Data?

POPGENE has the following main features and capacities that may be of interest to you during its use for your data analysis:

Windows: In view of potentially large raw data sets and outputs encountered in population genetics analysis, we have spent a considerable amount of time and effort to develop a Window text file editor that enables a loading/display of **very large** documents (Much of baseline information needed to develop our Window text editor was found in BIGED, the FREE source code distributed by Terry Richards Software). This editor is *not* limited to 64K of text, a drawback of many Windows applications written in Borland C++. Now, with POPGENE, the only limitation to the size of your data is how much memory (RAM) your computer has!

Other convenient and useful features that we have developed are assembled in the **Tool Bar**. You may change fonts and font sizes of your document so that you can, for example, view very long lines in the document. You may also cut and paste your output for inclusion into your word processing software.

Computing: The modules for co-dominant and dominant markers are currently limited to a maximum of:

1400 populations;
150 groups;
1000 loci;

10 characters (Alpha-numeric) for a locus name (automatically truncates to 10 if more than 10 characters are given).

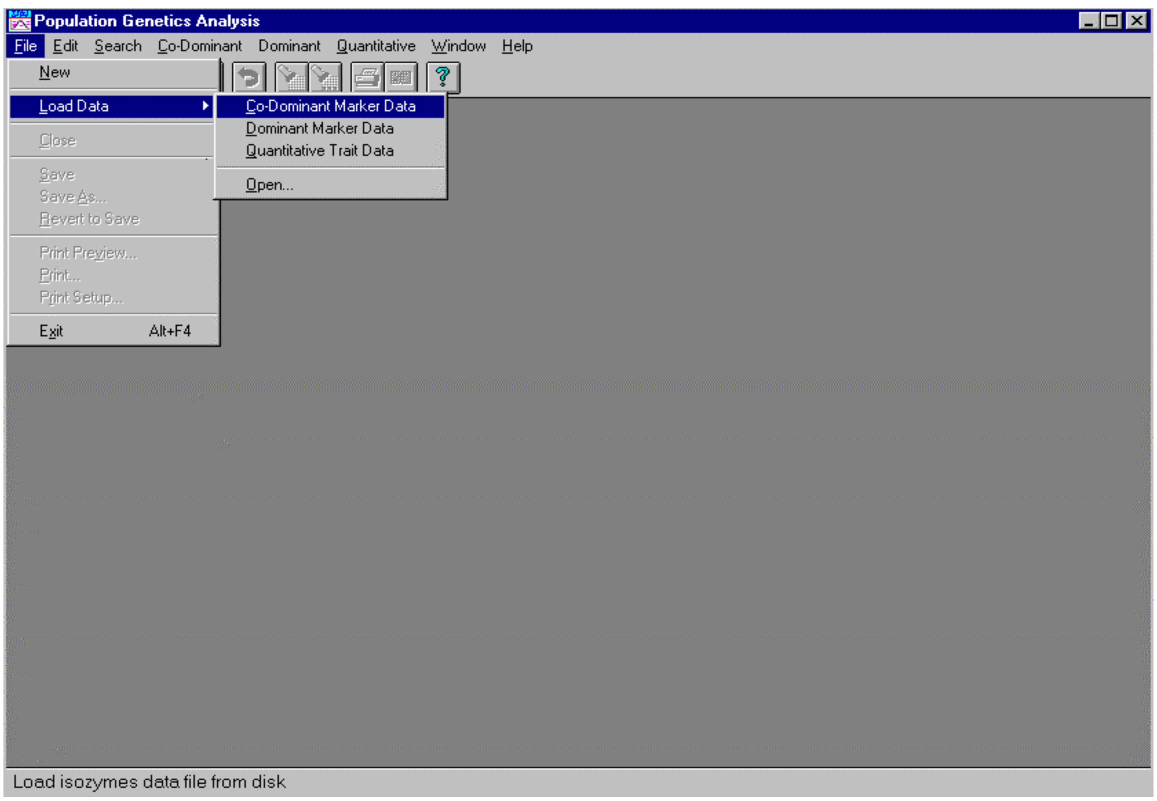
The number of alleles per locus is limited to 9 (**1-9**) if you use the numerals to code your alleles or to 52 if you use the alphabetic letters (respectively, capital alphabet **‘A - Z‘** for alleles **1 to 26** and lower alphabet **‘a -z‘** for alleles **27-52**).

Getting Started

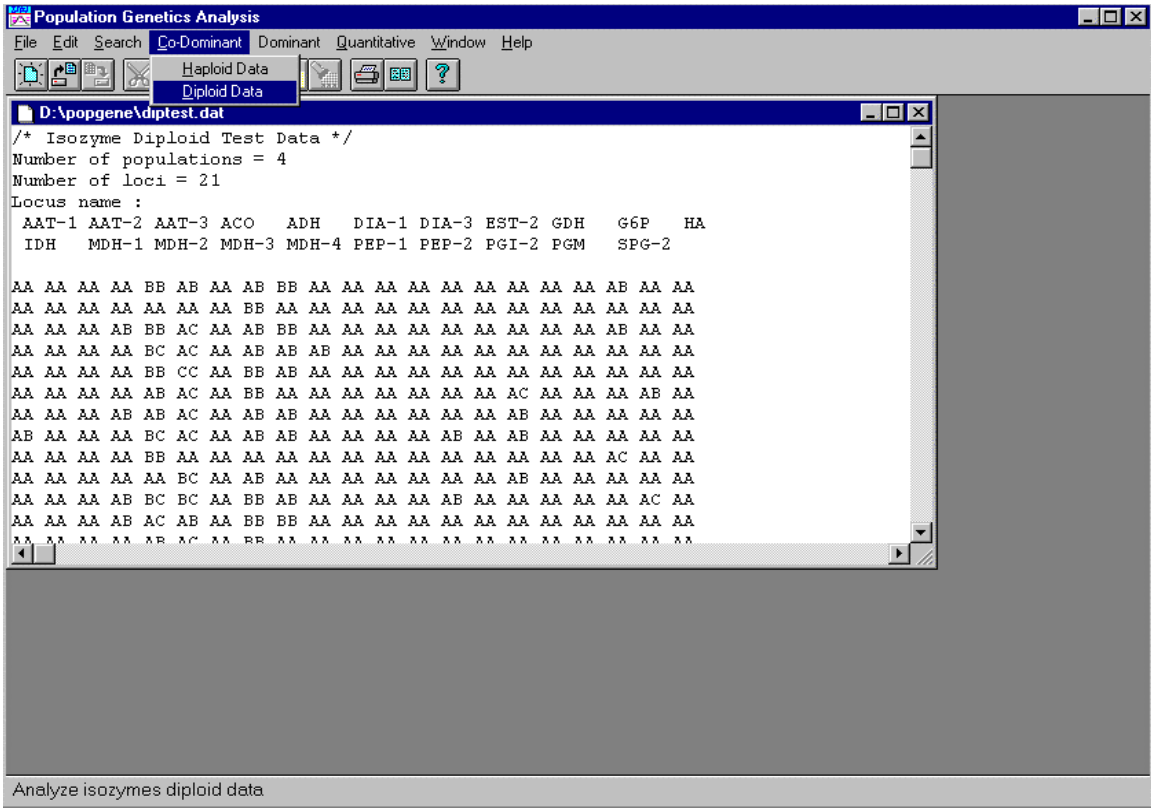
Before being able to conduct any analysis with **POPGENE**, you will have to prepare an ASCII input file (see **Input File Format** for different formats required for your input data files). Any commonly used text editor of your preference can be used for this purpose or you can bring data from other Window packages (*e.g.*, Microsoft Excel) into the **POPGENE** text editor, using **cut-and-paste**. Alternatively, you can use **POPGENE** Window-based text editor to prepare the required data file. ***Save your file before you proceed further!***

To carry out the analysis with **POPGENE**, proceed with the following steps:

(1) Click on **File** on the main menu bar and **Load Data** on the **File** menu to select appropriate data sets to be analyzed (in this version, **Co-Dominant and Dominant Data** can be selected for further analysis).



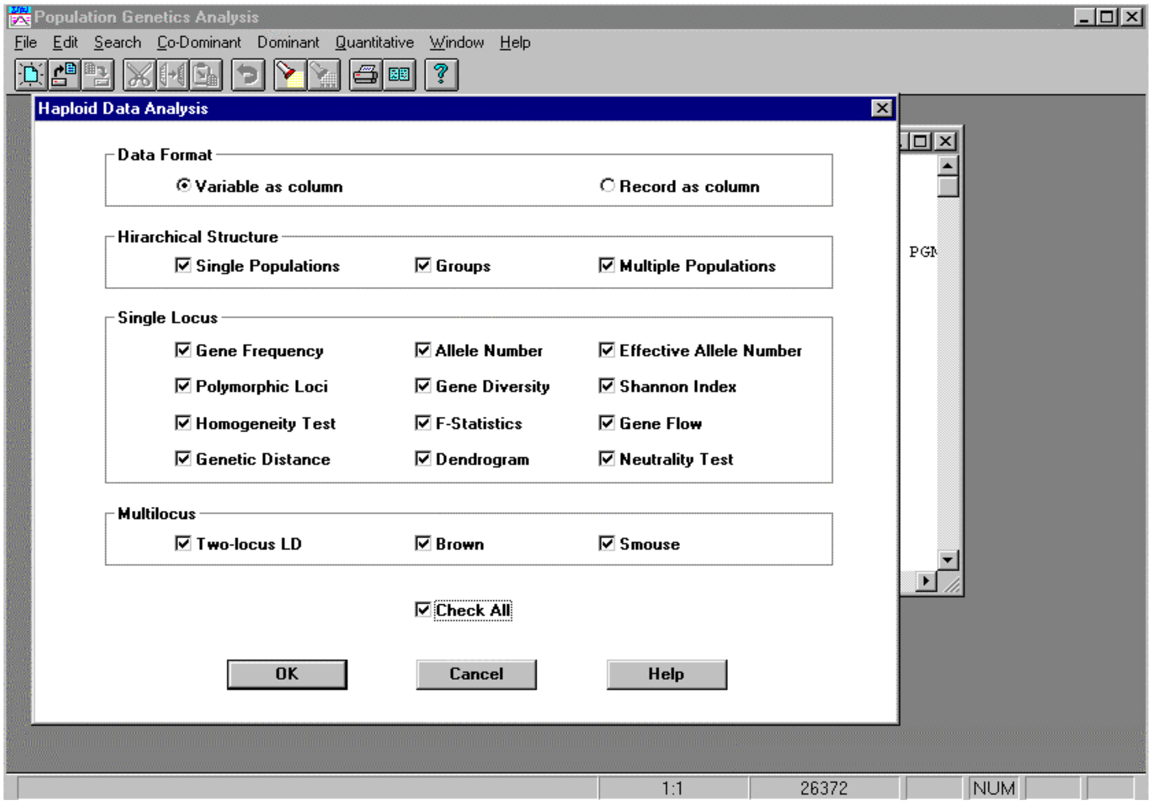
(2) Upon loading your data file, click on **Co-Dominant** on the main menu bar to select **Haploid** or **Diploid** analysis, depending on your data type.



(3) Open **Haploid Data Analysis** or **Diploid Data Analysis** dialog box to check:

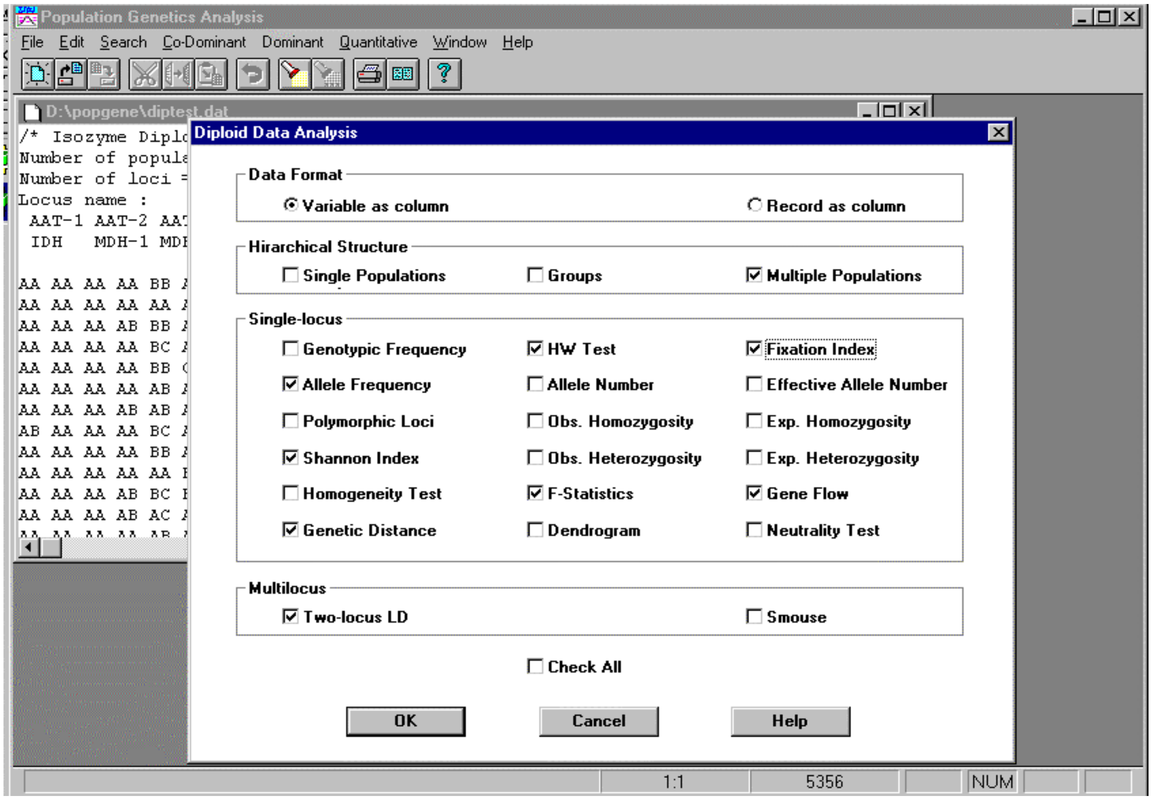
If variables (marker loci) or records (individual organisms) are entered as columns;
If your analysis will be carried out for **Single Populations**, **Groups** and/or **Multiple Populations**;
Appropriate **Single Locus** summary statistics;
Appropriate **Multilocus** summary statistics;

This is a **Haploid Data Analysis** dialog box. Notice that **Check All** was activated

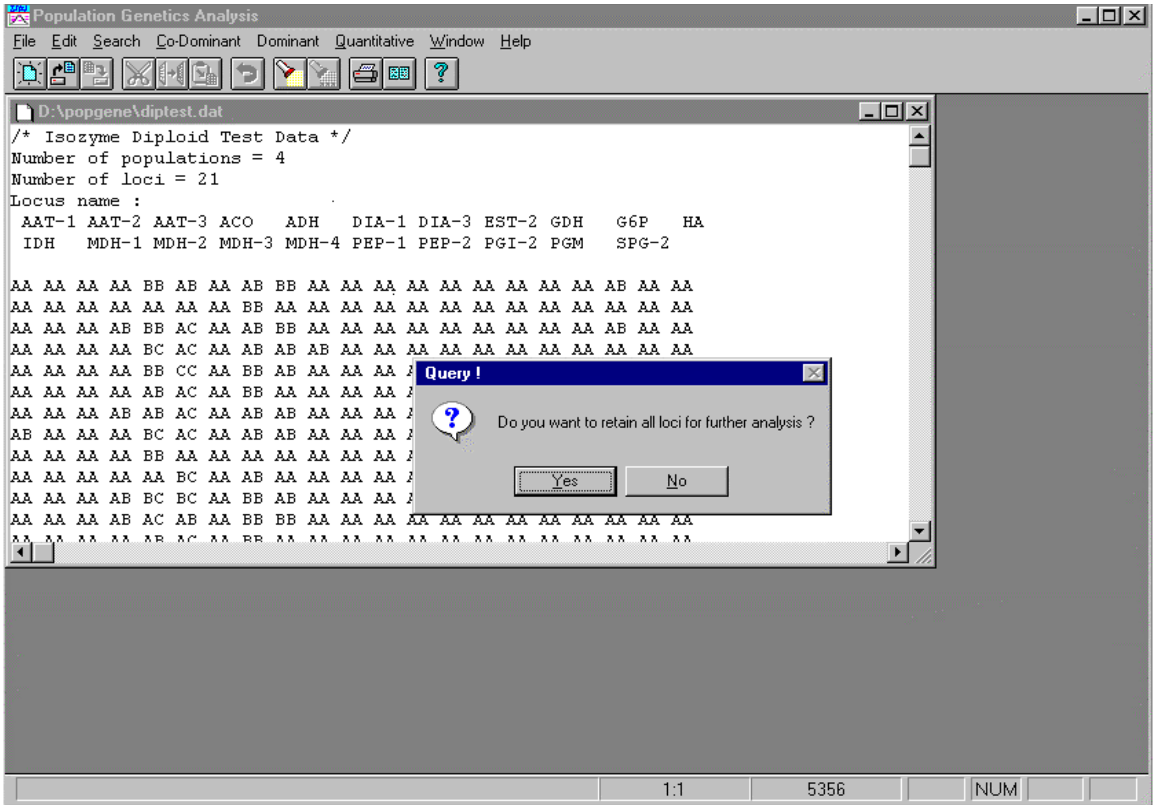


Check all if you are not sure what specific analysis you want to carry out and then click **OK**.

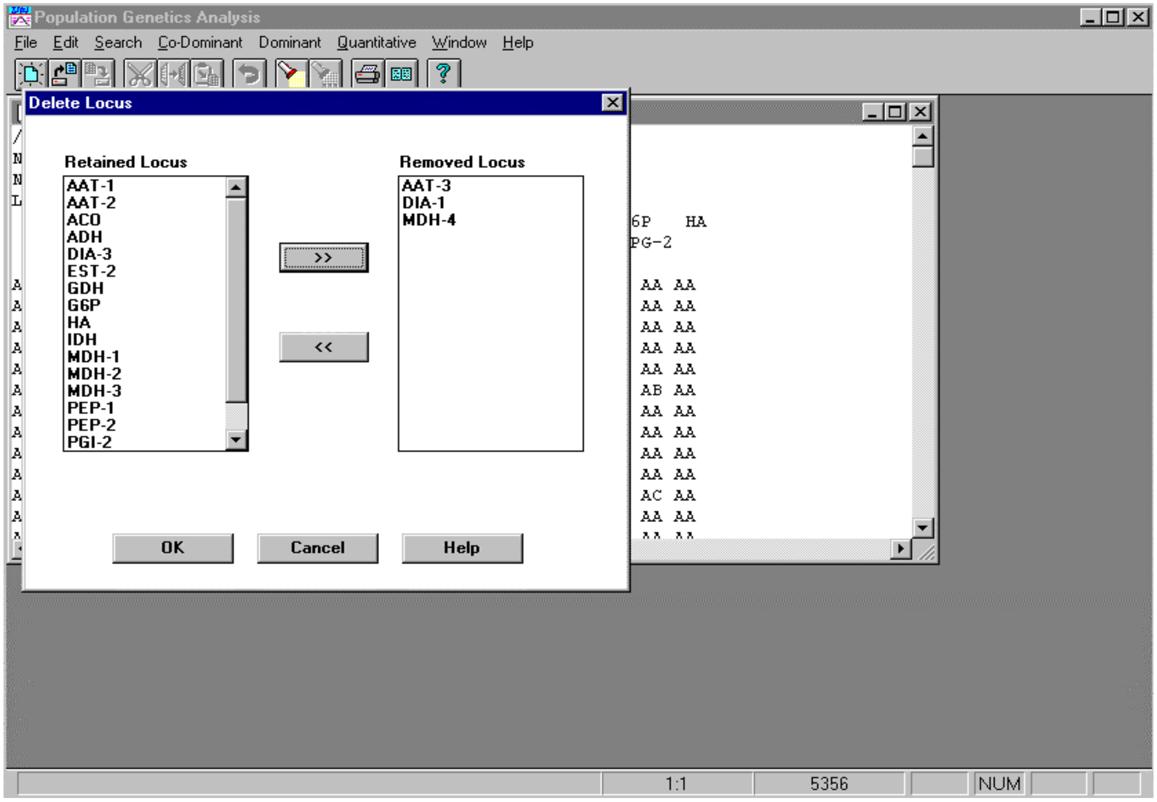
This is a **Diploid Data Analysis** dialog box for **co-dominant markers.** Notice that only some analyses were selected.



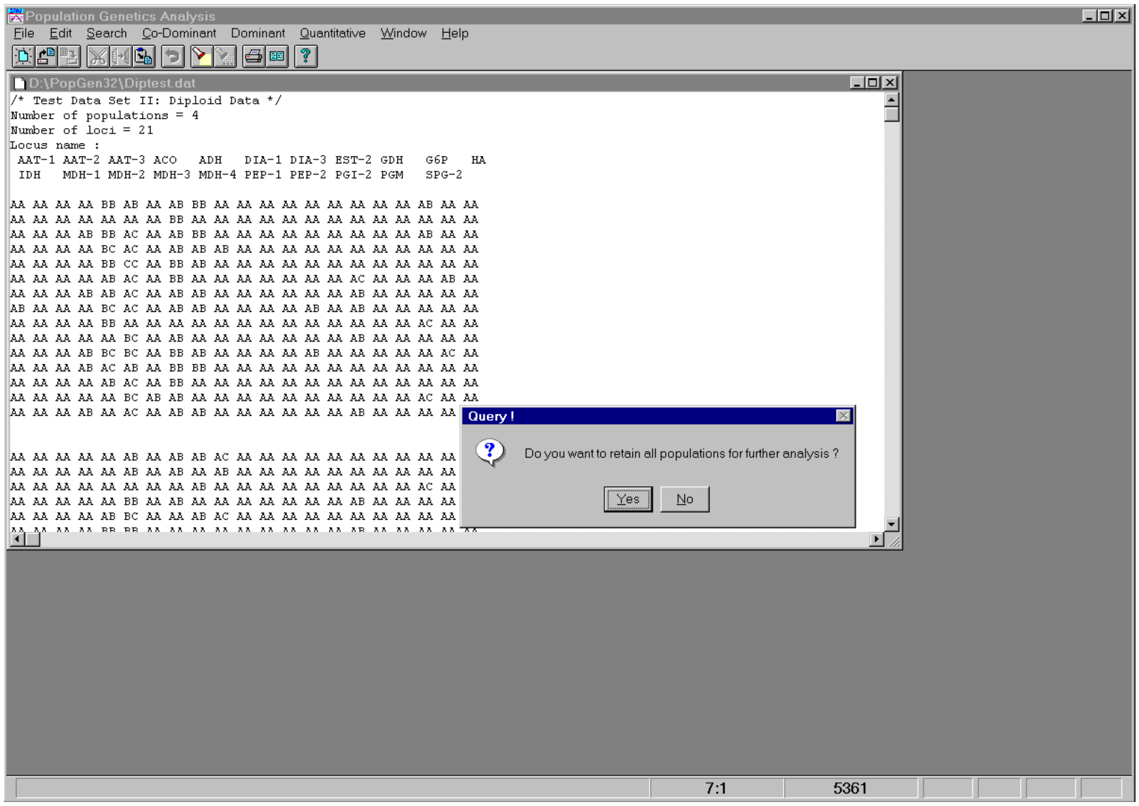
4) You are now prompted a Query "Do you want to retain all loci for further analysis ?" Click on **Yes** or **No** to answer the question.



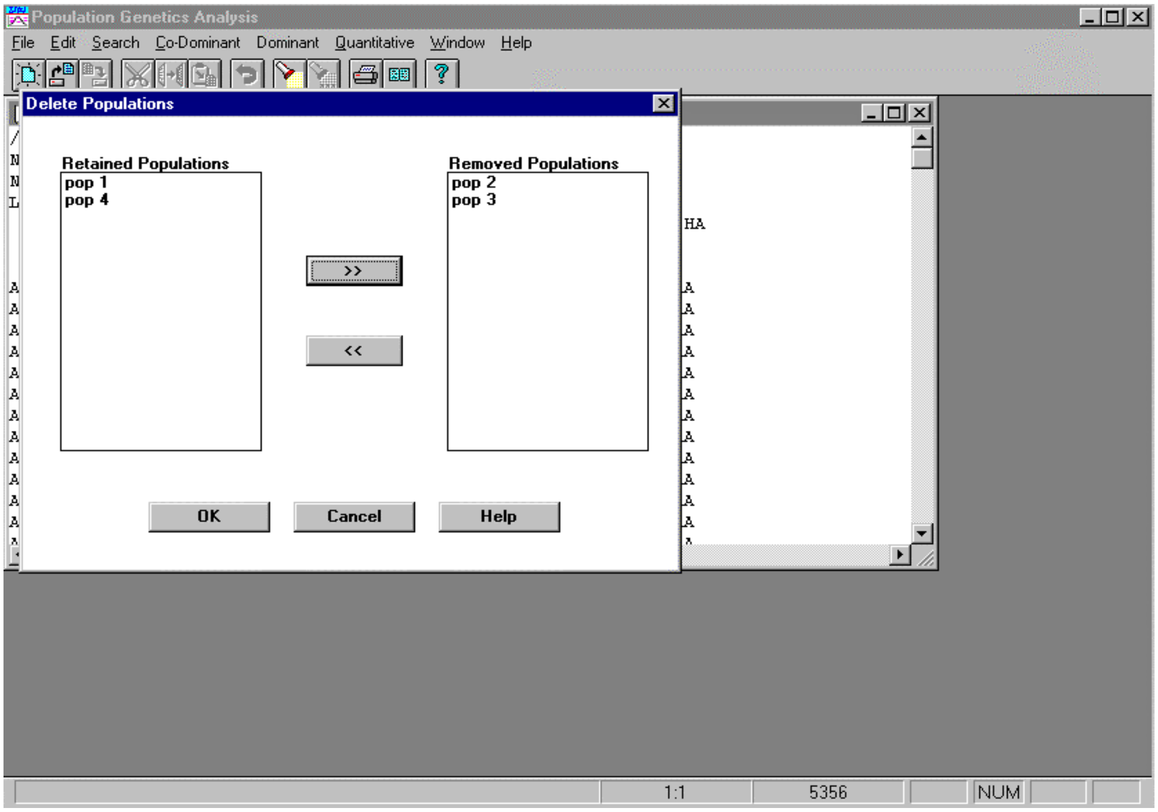
If your answer is **No**, then **Delete Locus** dialog box pops up for your selection of loci to be deleted.



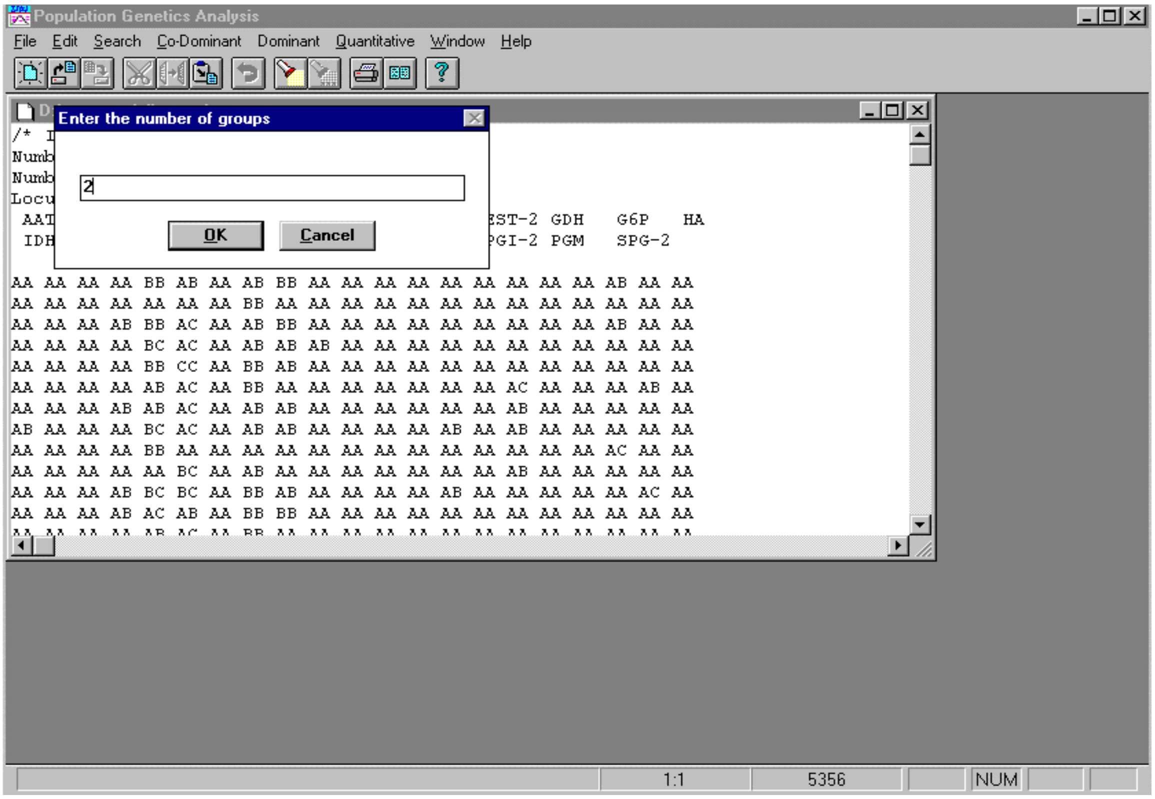
5) Next, answer the question "Do you want to retain all populations for further analysis?



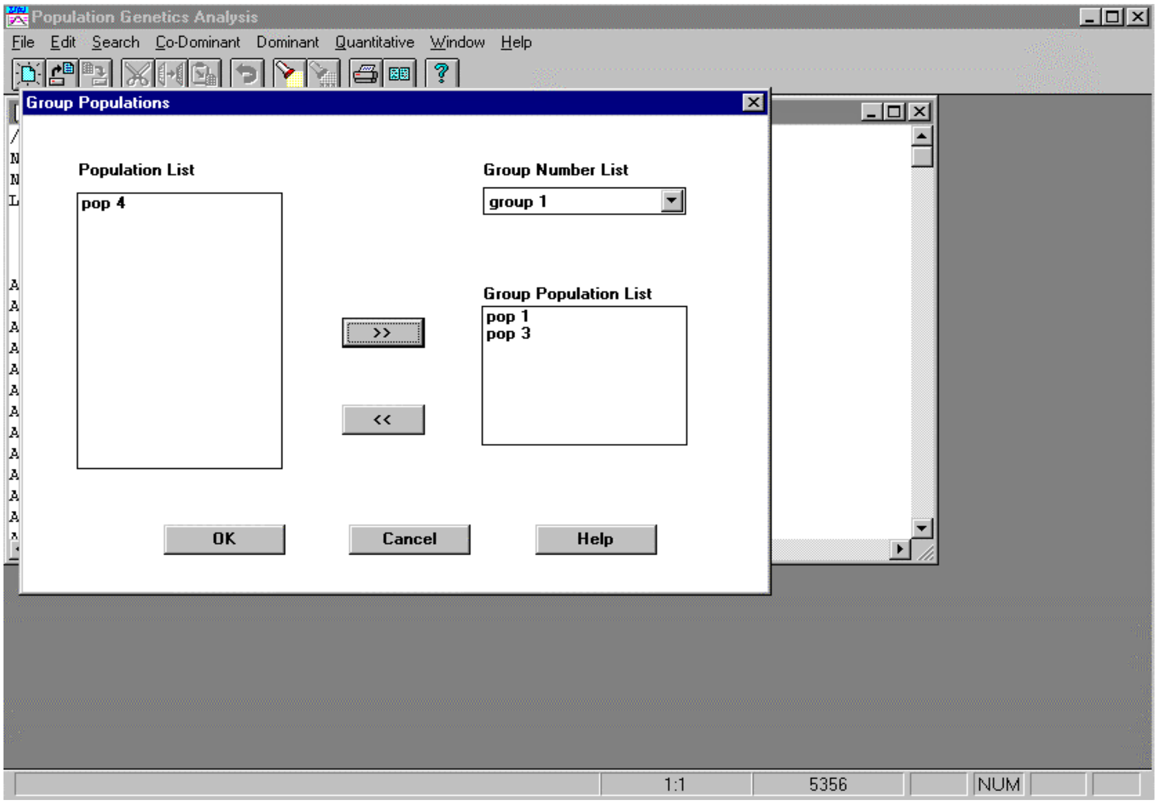
If your answer is **No**, then **Delete Populations** dialog box pops up for your selection of populations to be deleted.



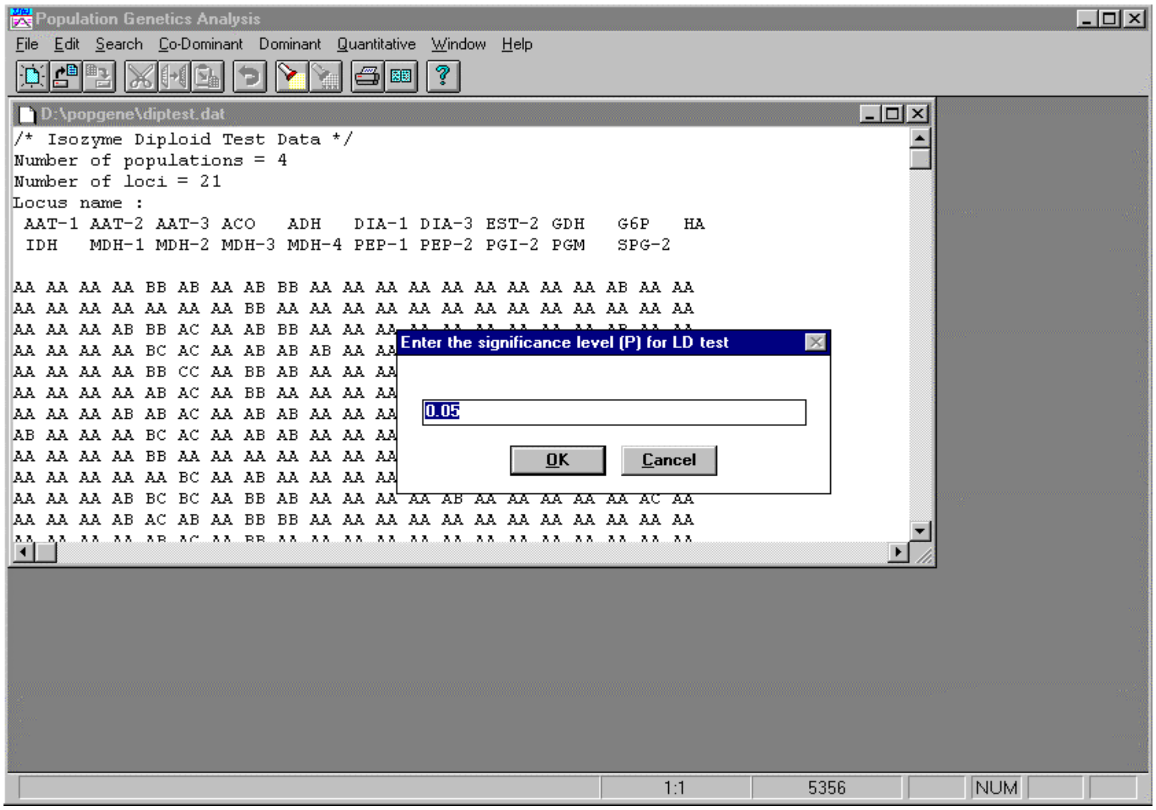
6) If you selected **Groups** at step (3), enter the number of groups in **enter the number of groups** dialog box;



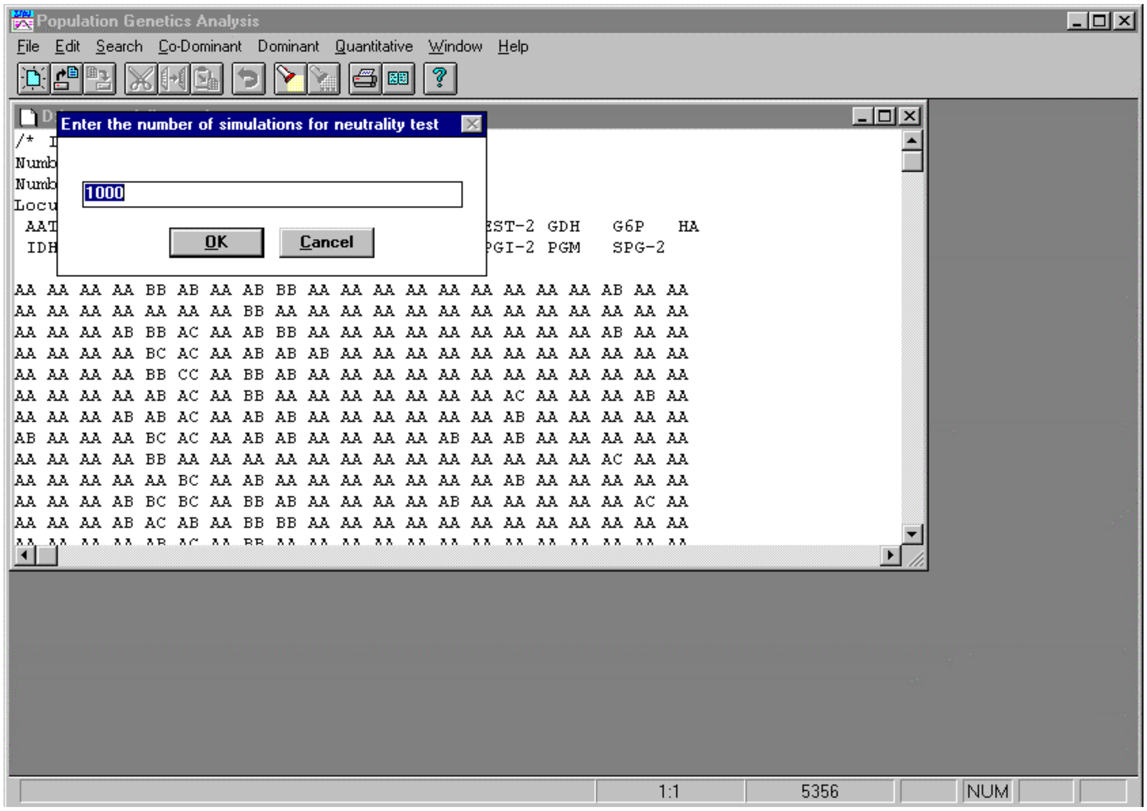
Click **OK** to open **Group Populations** dialog box for grouping of populations and select appropriate populations for each group.



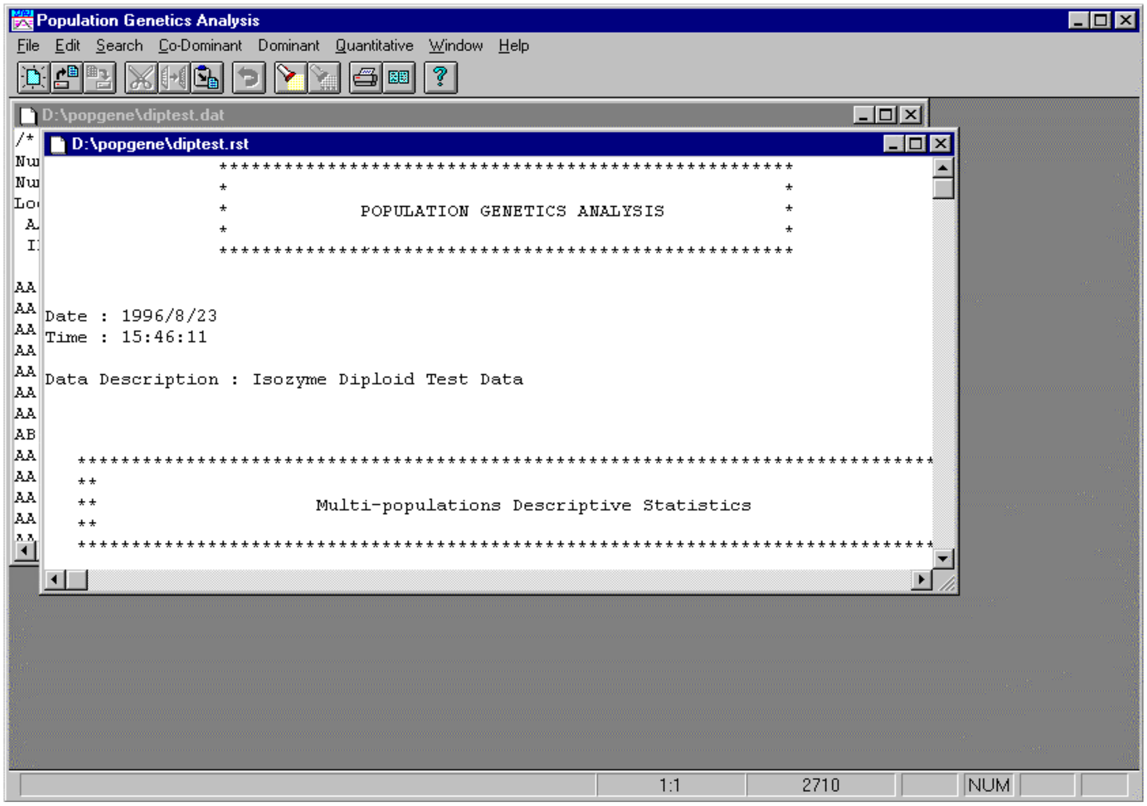
(7) If **Two-locus LD** was checked at step (3), then you need to select a significance level (P) to test for linkage disequilibria between pairs of loci. **Important: A high P value can result in an extremely large output when you have a large number of alleles/locus, loci and populations!!!** In most cases, *P* £ 0.05 should be used.



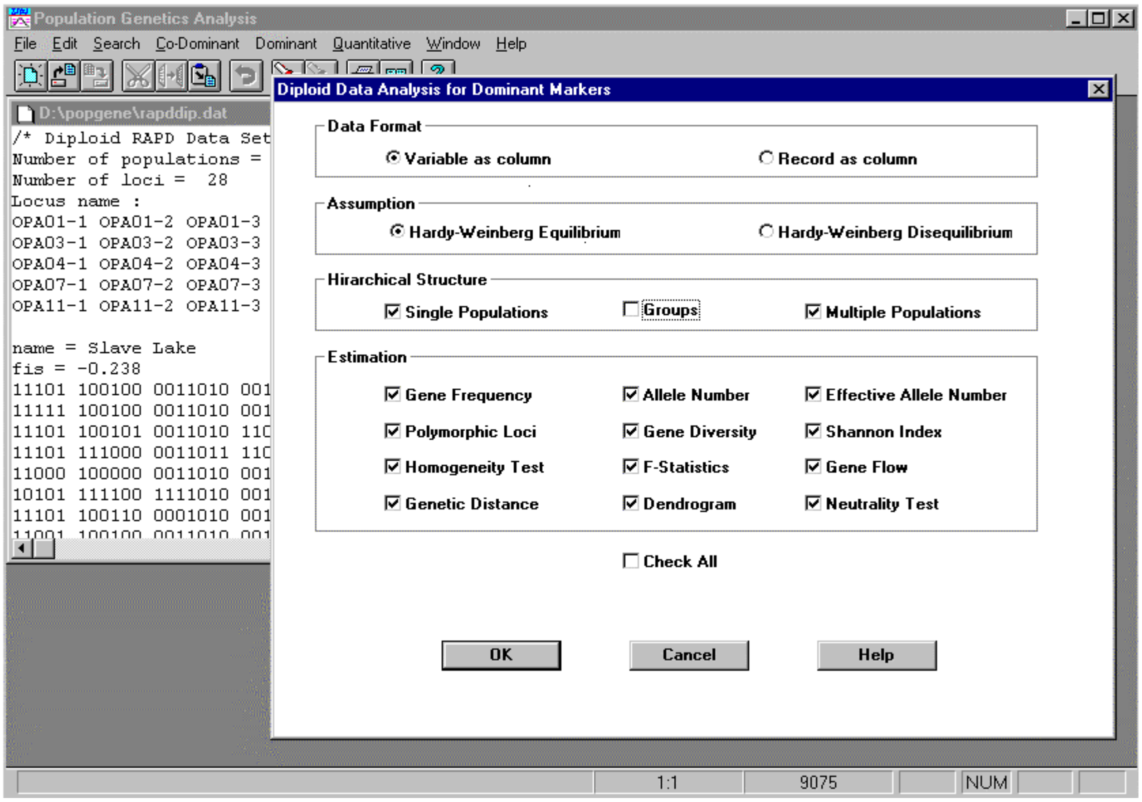
(8) If **Neutrality Test** was checked at step (3), then you need to select the number of simulations for computing 95% lower and upper confidence limits used to test for neutrality. We recommend 500 - 1000 simulations for a reliable estimation of these confidence limits.



(9) If you have correctly completed steps (1) to (8), then you are prompted with **result.dat** output Window, displaying the results from the analysis you just chose. Activate the menu **File | Save as...** to save the output into an ASCII file for further use or cut-and-paste selected text directly into a Window-based word processing package, such as Microsoft Word or WordPerfect.



This is a **Diploid Data Analysis** dialog box for **dominant markers.** You are prompted to choose between HW equilibrium (*i.e.*, F_{IS} = 0) or HW disequilibrium (*i.e.*, $F_{IS} \neq 0$). Notice that in your data, you can specify F_{IS} value for each population. If you do not have an input for F_{IS}, the program will assume that the population is in HW equilibrium. When you select HW disequilibrium at the prompt, the program will read your F_{IS} value and use it when estimating allele frequency. If you have F_{IS} value for each population but select HW equilibrium, the program will ignore your F_{IS} value and assume HW equilibrium.



All other aspects of data analysis for **dominant markers** are similar to that described for **co-dominant markers**.

Input File Format

Input file for **POPGENE** analysis consists of the header section and the data. The header section specifies (1) a job title delimited by **/* ... */;** (2) number of populations; (3) number of loci and (4) locus names. The body of data starts with, for each population, population ID # (optional), population name (optional). If you do not give population ID # or population name, you must leave at least one blank line between populations and **POPGENE** will generate population ID # automatically for you. But if you do, your population ID # and population name must be unique for each population. The raw data, in free format and ***with or without one or more spaces between columns***, immediately follow *without blank lines in between*. Missing values must be set to " . ‘‘ for haploids and dominant markers such as RAPDs (*i.e*., one digit to score for presence or absence of allele) and ".." (*i.e*., two digits) for diploids co-dominant markers in your input file. Here are three examples of data. The first two datasets have space between columns and the third dataset has mixture of space and without space to illustrate flexibility of data input and how input file for haploid and diploid should be prepared.

<u>Input file format for haploid data</u>

```
/* Haploid numeric data of 3 populations each with 3 records (gametes) & 19 loci
*/
Number of populations = 3
Number of loci = 19
locus name :
AAT-1 AAT-2 ACO ADH APH DIA-2 DIA-3 GDH G6H IDH MDH-1 MDH-2 MDH-3 MDH-4 ME PGI
PGM 6PG-1 6PG-2

ID =    1
 1  1  1  2  1  1  1  1  2  1  1  1  2  1  1  1  1  1  1
 1  1  1  2  1  1  1  1  2  1  1  1  2  1  1  1  1  1  1
 3  1  1  3  1  1  1  1  2  1  1  1  2  1  1  1  2  1  1

ID =    2
 1  1  1  2  1  1  1  1  2  1  1  1  3  1  1  1  1  1  1
 1  1  1  1  1  1  1  1  2  1  1  1  6  3  1  1  1  1  1
 1  1  1  1  1  1  1  1  2  1  1  3  3  1  1  1  1  1  1

ID =    3
 1  1  3  2  3  1  1  1  1  1  1  1  2  1  1  1  1  1  1
 1  1  1  2  3  1  1  1  2  1  1  1  2  1  1  1  1  1  1
 1  1  3  2  1  1  1  1  1  1  1  1  2  1  1  1  1  1  1
```

Input file format for diploid data, co-dominant marker

```
/* Diploid alphabetic data of 3 populations each with varying records
(genotypes) & 21 loci */
Number of populations = 3
Number of loci  = 21
Locus name :
 AAT-1 AAT-2 AAT-3 ACO   ADH   DIA-1 DIA-3 EST-2 GDH   G6P   HA
 IDH   MDH-1 MDH-2 MDH-3 MDH-4 PEP-1 PEP-2 PGI-2 PGM   SPG-2

AA AA AA AA AA AA AA BB AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AB BB A3 AA AB BB AA AA AA AA AA AA AA AA AA AB AA AA
AA AA AA AA BC AC AA AB AB AB AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA BB CC AA BB AB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AB AC AA BB AA AA AA AA AA AA AA AC AA AA AA AB AA
AA AA AA AB AB AC AA AB AB AA AA AA AA AA AA AB AA AA AA AA AA
AB AA AA AA BC AC AA AB AB AA AA AA AA AB AA AB AA AA AA AA AA
AA AA AA AA BB AA AA AA AA AA AA AA AA AA AA AA AA AA AC AA AA
AA AA AA AA AA BC AA AB AA AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AB BC BC AA BB AB AA AA AA AA AB AA AA AA AA AA AC AA
AA AA AA AB AC AB AA BB BB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AB AC AA BB AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AA BC AB AB AA AA AA AA AA AA AA AA AA AA AC AA AA
AA AA AA AB AA AC AA AB AB AA AA AA AA AA AA AB AA AA AA AA AA


AA AA AA AA BB BB AA AA AA AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AA AC AC AA BB AB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AB BB BC AA BB AA AC AA AA AA AA AA AA AA AA AC AD AA
AA AA AA AA AB BC AA AB AB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA BB BC AA BB AA AA AA AA AA AA AA AB AA AA AA AC AA
AA AA AA AA AA BC AA AB AB AA AA AA AA AA AA BC AA AA AA AA AA
AA AA AA AA AB BC AA AB AB AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AA BD BC AA AB AB AA AA AA AA AA AA AA AA AA AC AE AA
AA AA AA AB CC BB AA BB AA AA AA AA AA AA AA AA AA AA AA AB AA
AA AA AA AA BC BB AA AB AB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AB CC BB AA AA AB AA AA AA AA AA AA AA AA AA AA AB AA
AA AA AA AA BB BC AA AB AB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA AB BC AA AA BB AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AB AC AB AA BB AA AA AA AA AA AA AA AA AA AA AA AB AA
AA AA AA AB BC BB AA BB BB AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AB AB AB AA BB AA AA AA AA AA AA AA AA AA AA AA AA AA


AA AA AA AA BB AB AA AA AA AA AA AA AA AA AA BB AA AA AA AA AA
AA AA AA AA AC BC AA BB AA AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AA CC BC AA AB AB AA AA AA AA AA AA AB AA AA AA AC AA
AA AA AA AA BB AC AA BB AA AA AA AA AA AA AA AA AA AA CC AA AA
AA AA AB AA BE BB AA BB AB AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AA AB BC AA AB AA AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA BB AA CC AA AB AB AA AA AA AA AA AA BC AA AA AA AA AA
AA AA AA AB AC BC AA BB BB AA AA AA AA AA AA AA AA AA AA AA AA
AA AA AA AB AA BB AA AC AB AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AB BB AC AA BB AA AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AA AC AC AA AA AB AA AA AA AB AA AA AB AA AA AA AA AA
AA AA AA AA AA AB AA AB AB AA AA AA AA AA AA AB AA AA AA AA AA
AA AA AA AA BC BC AA BB AB AA AA AA AA AA AA AB AA AA AA AC AA
```

Input file format for diploid data, dominant marker

```
/* Diploid RAPD Data Set */
Number of populations =  2
Number of loci =  28
Locus name :
OPA01-1 OPA01-2 OPA01-3 OPA01-4 OPA01-5
OPA03-1 OPA03-2 OPA03-3 OPA03-4 OPA03-5 OPA03-6
OPA04-1 OPA04-2 OPA04-3 OPA04-4 OPA04-5 OPA04-6 OPA04-7
OPA07-1 OPA07-2 OPA07-3 OPA07-4 OPA07-5 OPA07-6
OPA11-1 OPA11-2 OPA11-3 OPA11-4

name = Slave Lake
fis = -0.238
11101 100100 0111010 001000 1001
10110 100100 0011010 001000 0001
11100 100100 0011010 001000 0001
11111 100100 0011010 001010 0001
10110 101100 0011010 001010 1101
11000 100100 1001010 001000 0001
11101 100100 0101010 001000 1001
10110 100100 0000110 111101 0001
11001 100100 0111010 001010 1001
11101 101100 0101010 001000 1001
11100 100100 0011010 110000 1001
11101 111100 0011010 001000 0001
11111 111000 0101010 001000 0001
11000 100100 0000010 001010 0001
11001 100100 0110010 001000 1001
11101 101000 1000010 001000 1001

name = Little Smoky
fis = 0.0
11101 111101 0111010 111000 1111
11101 111100 0011010 001000 1111
11110 100001 1011010 101010 1011
11101 111000 0011010 001011 0001
10101 111111 0011010 111000 0101
11000 100000 0111010 111010 0111
10001 101000 0011010 011000 1001
11101 101111 0011010 011010 1001
11001 101010 0011010 101010 1011
11011 111100 0011010 011010 1011
11110 111101 0011010 001001 0011
11001 111111 1011010 001010 1001
11111 111110 0000010 111010 0111
11011 101000 0011010 001000 0001
11101 100000 1111010 001000 1001
11111 100000 1111010 001000 1001
11100 101100 0001010 000101 0110
11001 101100 0100010 110100 1001
10101 100100 0011010 000010 1001
11001 111000 0010010 001000 0001
```

How to Contact Us

While we have carried out a fairly extensive testing of the package using various data, we would appreciate that you let us know any bug or abnormal behavior of the program during your use with **POPGENE**. We would also like to hear from you on improvements required in future releases. POPGENE is on website at:

http://www.ualberta.ca/~fyeh/

or contact:

Dr. Francis C. Yeh Department of Renewable Resources
University of Alberta
Edmonton, AB Canada T6G 2H1
Tel.: (780) 492-3902, Fax: (780) 492-4323
Email: francis.yeh@ualberta.ca
Home Page: http://www.ualberta.ca/~fyeh/fyeh

Credits

Dr. Francis C. Yeh
Dr. Timothy Boyle
Dr. Yang Rongcai
Zhihong Ye
Judy Mao Xiyan

Acknowledgements

POPGENE cannot be initiated without financial support. We thank the Natural Sciences and Engineering Council of Canada for an operating grant (A2282) to FCY for population genetics research, the Canadian Forestry Service for a Green Plan Grant to FCY for studying multilocus genetic structure, and the Centre for International Forestry for a grant to FCY to support the analysis of multilocus variation, C++ programming, and Microsoft® Window interface.